Snapshot of Statistical Methods Used in Geriatric Cohort Studies: How Do We Treat Missing Data in Publications?

Diklah Geva[1,*], Danit Shahar[1], Tamara Harris[2], Sigal Tepper[1], Geert Molenberghs[3] and Michael Friger[1]

[1]*Department of Epidemiology and Health Science Evaluation, Faculty of Health Science, Ben Gurion University of the Negev, P.O. Box 653 Beer-Sheva 84105, Israel*

[2]*Laboratory of Epidemiology, Demography, and Biometry, Gateway Building, 3C309, 7201 Wisconsin Avenue, Bethesda, MD 20892, USA*

[3]*Center for Statistics (CenStat), Universiteit Hasselt, Agoralaan 1, B-3590 Diepenbeek, Belgium*

Abstract: *Background*: Geriatric studies often miss data of frail participants. The aim of this paper is to explore which missing data methodologies have entered current practice and to discuss the potential impact of ignoring the issue.

Methods: A Sample of 103 articles was drawn from key cohort studies: *Health ABC, InCHIANTI, LASA, BLSA, EPESE, and KLoSHA.* The studies were classified according to missing data methodologies used.

Results: Seventy-seven percent described the selected analysis data set and only 28% used a method of handling all available observations per case. Missing data dedicated methods were rare (< 10%), applying single or multiple imputations for baseline variables. Studies with longer follow-up periods more often employed longitudinal analysis methodologies.

Conclusions: Despite the recognition that missing data is a major problem in studies of older persons, few published studies account for missing data using limited methodologies; this could affect the validity of study conclusions. We propose researchers apply Joint Modeling of longitudinal and time-to-event data, using shared-parameter model.

Keywords: Missing data, geriatric cohort studies, methodologies review, longitudinal analysis.

INTRODUCTION

The population of people age over 60 is growing faster than any other age group worldwide. According to the World Health Organization (WHO) in 2007, the world's population in this age group was 650 million and by 2050 the aging population is forecasted to reach 2 billion. Extensive research is being carried out to better understand the epidemiology and biology of aging and the prevention and management of age-associated chronic diseases alongside the identification of factors that might help older people retain their health and remain a resource for society. Geriatric surveys and cohort studies are the main sources of evidence to support research of this age group. However, due to frailty of the older population, the data collected are often incomplete and favor the relatively more sturdy older participants. The generalizability of the findings published from these cohort studies largely depends on dataset completeness and the ability to account for missing data in the analysis. Ignoring the incomplete data may result in biased findings that could affect the validity of study conclusions.

Three statistical approaches are common for data analysis of geriatric studies: cross-sectional, survival, and decline curves analyses. Cross-sectional analysis is common in survey studies and it is used to obtain correlates of environmental factors with illness and functional indicators. The correlation estimates may suffer from selection bias when completers have better health than those who fail to complete all survey parts. Studies involving follow up may use survival analysis to describe the effect of different factors on accumulation of health events and deaths. Survival analysis offers an analytical framework where attrition, censoring, and deaths are integral parts of the data analytic setting and the estimation process. The generalizability of conclusions drawn from survival analysis depends on the correct classification of the cases that have been lost to follow-up or have missing data. Decline curve analysis is unique in its capacity to describe the effect of different factors on the deterioration process of a qualitative scale such as cognitive performance or walking speed. Likelihood-based analysis, such as done for the Linear Mixed Models (LMM) and Generalized Estimating Equations (GEE), would be a natural analysis framework for capturing the decline curves and testing for factors that may be associated with the decline. Nevertheless, in this analytical framework selection bias due to attrition and missing

*Address correspondence to this author at the Department of Epidemiology Faculty of Health Science, Ben Gurion University of the Negev, P.O. Box 653 Beer-Sheva 84105, Israel; Tel: ++ 972 52-3420020; Fax: ++ 972 153-3-6436295; E-mail: gevadi@post.bgu.ac.il

E-ISSN: 1929-6029/13

data also grows with follow up and requires analytical attention.

In recent years, tools for handling missing data have become available in most common statistical software packages. These tools are based on the important theoretical developments that took place over the past three decades. Growing concern about the prevention and treatment of missing data in clinical studies led to the development of guidelines by a panel of experts [1]. The same concerns are pertinent to epidemiological studies and a complete chapter was devoted to missing data in the Handbook of Epidemiology [2]. The aim of this paper is to explore how these promoted methodologies for handling missing data have penetrated current publications and to illustrate the potential impact of ignoring the issue. To this end, we obtained a snapshot of missing data methodologies in recent journal articles published from six key geriatric cohort studies.

METHODS

Review of publications from six key geriatric studies was carried out by the first author of this paper. All six studies are of older adults; however they varied greatly in terms of geography (Europe, USA, and Asia), primary biomedical focus, study design, and years and frequency of follow-up visits. Below is a brief description of the studies.

Health Aging and Body Composition (HealthABC) study was designed to assess body composition and physical functioning changes in older adults. It is a community-based study of n=3075 high functioning men and women aged 70-79 years at enrollment. Participants were recruited in 1997-1998 at two centers in the United States, Pittsburgh, PA, and Memphis, TN.

The InCHIANTI Study (InChianti) was designed to identify risk factors for late-life. It is a population-based study of n=1154 participants who were invited to the study. The sample was recruited during 1998-2003 and included older adults age 65+ from Greve, Chianti, and Bagno a Ripoli, Tuscany, Italy.

Longitudinal Aging Study Amsterdam (LASA) was designed to focus on physical, emotional, cognitive, and social functioning late in life. It is a population-based study, recruited in a number of waves from the registry of 11 municipalities in the Netherlands in 1992-1993. There were n=3107 subjects, aged 55 to 85 years, who enrolled in the baseline phase of the first enrollment wave.

Baltimore Longitudinal Study of Aging (BLSA) is America's longest-running scientific study of human aging, begun in 1958. The study aim is to learn what happens as people age and how to sort out changes due to aging. More than 1,400 men and women are study volunteers. They range in age from their 20s to their 90s. Under the umbrella of this study, additional numerous studies were initiated.

Established Populations for Epidemiologic Studies of the Elderly (EPESE) are studies of the older population aimed at describing and identifying predictors of mortality, hospitalization, and placement in long-term care facilities, and to investigate risk factors for chronic diseases and loss of functioning. The original cohort began in the 1980s in East Boston, New Haven, and rural Iowa with a later addition from North Carolina centered at Duke (1993–1994) with n=3,050 Americans aged ≥65 years. An additional cohort begun in 1993, the Hispanic EPESE, included n=3050, aged ≥75 years, of community-based older non-institutionalized Mexican Americans residing in five southwestern states (Texas, California, Arizona, Colorado, and New Mexico).

Korean Longitudinal Study on Health and Aging (KLoSHA) was designed as a population-based prospective cohort study on health, aging, and common geriatric diseases of Korean elders aged 65 years and over. N=992 participants were recruited during 2005-2006 in Seongnam. This study had the shortest period of follow up at the time of this review and most publications, therefore, represent only study design and analysis of baseline characteristics.

Review of Publications

An NCBI PubMed (www.ncbi.nlm.nih.gov/pubmed) search conducted during August-September 2011. The initial list was generated using the study full name and its abbreviation. The limits activated in the PubMed search were: English language, Journal Article, humans, Aged: 65 years and over, and the paper had to be published in the last 2 years. Results that were not related to the specific study were omitted and the large lists in Health ABC and in LASA were limited to the first two dozen publications. The KloSHA study search period was extended to 3 years to allow for a larger number of publications. This generated a sample of n=103 relevant publications for review (see list according to studies and publication date in: https://sites.google.com/site/diklahgeva/). The selected publications were examined and the attributes that

were summarized into the database included: first author name, journal and publication date, title of study and information about dataset selection, statistical methods used in the research, and missing data methods used in data selection and in the analysis such as the use of complete case (CC) versus available observations of each case (AC). We particularly looked for dedicated methods including single imputations, comparison of completers to missing subgroups, and advanced methods such as multiple imputations, Selection Models, and Inverse Probability Weighted Regression either with single outcome or with repeated measures in a mixed model (MM) or generalized estimating equations (GEE) framework.

Classifications of the Publications According to Categories

After data collection, the papers were classified according to broad classes of analyses:

A. Cross-Sectional publications focus on ecological and environmental predictors of prevalence or development of new diagnostic index. The common statistical methodologies used in cross-sectional studies are correlations, contingency tables, regression, analysis of variance, and logistic regression.

B. Survival Analysis focuses on factors associated with event or death and uses life tables with Kaplan-Meier test and Cox Proportional Hazard Regression.

C. Decline Curves focused on capturing the mean decline over time and establishing cofactors affecting the downward process. Mixed-model methodology and Generalized Estimating Equations (GEE) are the main statistical tools used to estimate the mean decline curve.

The publications were also classified according to the outcome biomedical discipline:

Biomarkers including lab work and, genetic and biological markers intended to predict outcome; Function performance such as Daily Living Activities, walking ability and speed, balance, muscle strength, sensory assessments of hearing or vision;

Chronic disease including cardiovascular, metabolic, endocrine, diabetic, body composition and bone, in addition to cancer and neurology;

Psychiatry included depression, cognitive capacity, dementia, social factors, and psychoneurology functioning.

Public Health (PH) issues such as incidence, prevalence, trends, risk factors, demographic factors, mortality and morbidity, falls, trauma, development of diagnostic scale.

Life Style patterns including physical activity, nutrition, food supplements, minerals and vitamins, overweight and obesity.

Missing Data Classifications

All papers were indexed according to missing data treatment in three parts of the paper:

1) Data set selection: whether detailed description of cases that were included-to or omitted-from analysis due to missing data;

2) Methods: whether the author used all available observations from a case (AC) or whether it was a complete case analysis (CC);

3) Statistical analysis: whether dedicated missing data methods were used, for example single value imputation or comparison of baseline values of selected and omitted cases, in addition to the advanced methods such as multiple imputation, Inverse Probability Weighting (IPW) regression models or selection models; and whether any form of sensitivity analysis was carried out.

Summary Statistics

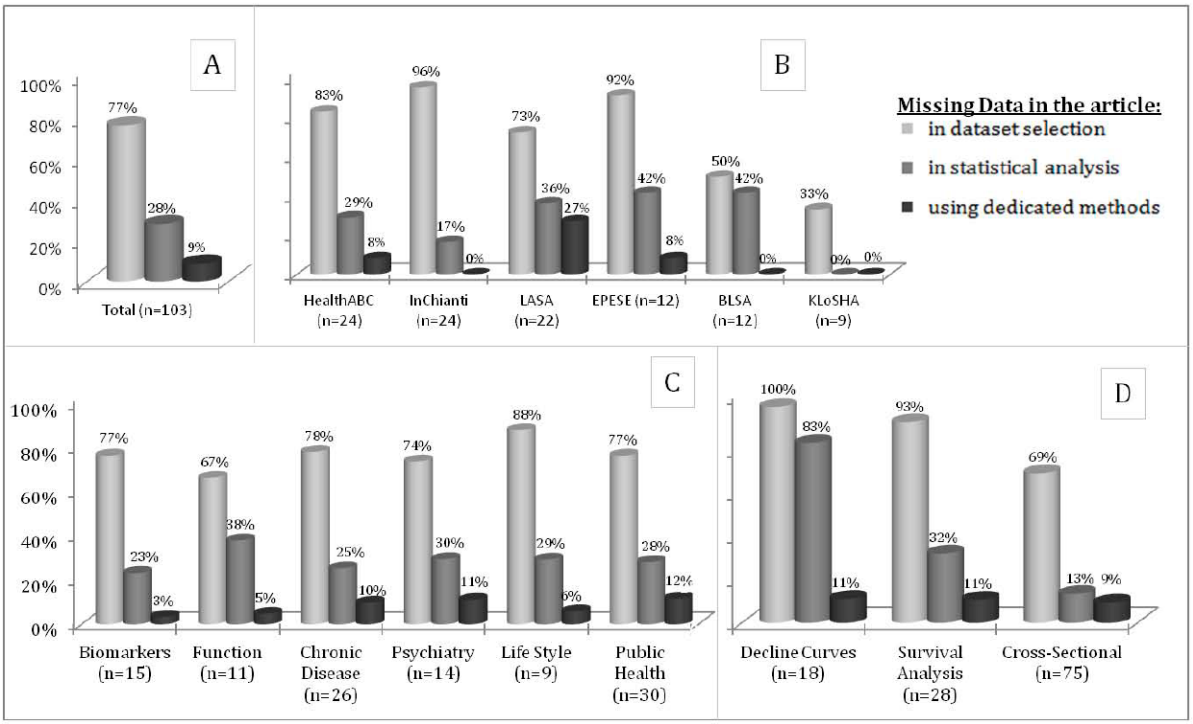
Counts, percentages, and cross-tabulations were used to describe the overall and categorical proportion of articles that included different types of missing data methodologies. Since most of the publications cover more than one area (e.g., lifestyle and chronic disease), all studies were classified into two possible areas – indices that were equally weighted in the cross-tabulation analysis of the biomedical areas. The results are presented in graphs of the percentages.

RESULTS

A sample of n=103 publications from six key studies of older populations was drawn. The selected publications were reviewed and classified according to biomedical area and statistical approach. See https://sites.google.com/site/diklahgeva/ for the

complete list. The publications were categorized according to handling of missing data in three aspects of the article: dataset selection, methods, and statistical analysis — whether a dedicated missing data method was employed. The findings from the review were tabulated and graphed. Figure **1** A shows that a total of 77% of the publications reported in detail on the data selection to analysis, yet only 28% used a statistical method for handling all available observations per case (AC) in the analysis. The other 72% of the publications used complete case (CC) analysis. Methods that are dedicated to missing data issues were rare, less than 10%; these few studies reported on the use of single imputation or conducted a comparison to the cases not included in the analysis. Less than a handful of the studies conducted multiple imputations with respect to baseline variables. None of the studies reported the use of advanced methods such as Inverse Probability Weighting (IPW) regression models or Selection Models to account for non-random incomplete data.

The studies were similar with respect to reports of missing data in the selection of the data set: 73-96% of publications in Health ABC, InChianti, LASA, and EPESE. Publications from two studies paid less attention to the missing data issues: the KloSHA, which is the most recent study, and BLSA, which is the oldest study (over 50 years), with many publications reporting



B. Studies:

Health ABC - Health Aging and Body Composition
InChianti - The InCHIANTI Study
LASA - Longitudinal Aging Study Amsterdam
EPESE - Established Populations for Epidemiologic Studies of the Elderly
BLSA - Baltimore Longitudinal Study of Aging
KLoSHA - Korean Longitudinal Study on Health and Aging

C. Biomedical areas:

Biomarkers - lab works, genetic, any biological marker intended to predict outcome.
Function - performance such as Daily Leaving Activities, walking speed, balance, muscle strength, hearing and vision.
Chronic disease - cardiovascular, metabolic, endocrine, diabetic, body composition and bone, cancer and relevant neurology.
Psychiatry - depression, cognitive capacity, dementia, social factors and related neurology.
Public Health - incidence, prevalence, trends, risk factors, demographic factors, mortality and morbidity, falls, trauma, diagnostic scale.
Life Style - physical activity and nutrition, food supplements, minerals and vitamins, overweight.

D. Statistical Methods:

Cross-sectional - ecological predictors of prevalence or development of new diagnostic index, commonly using correlation and regression analyses.
Survival analysis - factors associated with event or death; uses life tables and Cox Proportional Hazard Regression.
Decline Curves - drawing the decline over time and establish factors effect on the turn down process, Linear Mixed Models, Generalized Linear Models and Generalized Estimating Equations (GEE).
The percentages do not sum to 100% as papers may be classified on more than one category.

Figure 1: This figure shows percent of publications with missing data in 3 parts of the paper: Data-set selection (light gray), Statistical analysis methods (gray), and Dedicated missing data methods (dark gray). The first panel provides the percentage for All the publications (**A**), by Study (**B**), by Biomedical area (**C**), and by the underlying statistical method (**D**).

on the analysis of subsets of the data. Longitudinal studies with considerable periods of follow-up, such as Health ABC, LASA, EPESE, and BLSA, had 29-42% publications involving statistical methods that account for all available observations per case in the statistical model. LASA had the largest percentage, 27%, of publications that employed dedicated methods to deal with missing data issues, generally taking the form of comparison at baseline and single value imputation of baseline missing values of particular missing lab values.

The studies varied with respect to biomedical area: Health ABC had 31% publications on public health issues and 27% on chronic diseases; InChianti had 31% on chronic diseases and 25% on biomarkers; LASA had 39% on Public Health issues and 18% on chronic diseases; EPESE also focused on public health issues, 38%, and psychiatric conditions, 29%. BLSA had 21% biomarkers publications and 29% Public Health publications. Finally, KloSHA had 33% Public Health Issues, 22% chronic disease, and the same percentage on psychiatric conditions.

Figure **1C** displays the missing data issues by the biomedical area of the publication. Despite the very different nature of the biomedical area, the missing data handling is similar across areas; 67-88% of papers regard dataset selection, 23-38% of papers regard missing data in the statistical analysis methodology, and 3-12% of papers have dedicated methods. In contrast, Figure **1D** shows that the missing data profile is very different depending on the type of statistical approach. The majority of the publications (73%) had a cross-sectional analysis with the least attention to missing data, while decline curves analysis had the most regard to missing data, both in describing cases included in the analysis and in using a method that allows all available observations per case to be included in the analysis. Survival analysis had 32% of papers including missing data in the analysis, while decline curves analysis had more than double, 83%. In this review we did not classify the survival analysis method as providing an analytical tool for missing data, leading to the difference between 32% and 83%. Survival analysis is a common statistical method applied for the analysis of cohort data; it utilizes vital status indication while ignoring quantitative values, which are more frequently missing. While there is similarity between the issues dealt with by decline curves analysis and those occurring in survival data, the latter also have their peculiarities, which are not accounted for in the framework of this paper.

Some of the above differences may also be explained by the fact that the studies varied in terms of their length of follow up and thus their statistical approach. Although most of the papers had cross–sectional issues, the studies with longer periods of follow up more frequently used decline curves and survival. Health ABC had 42% survival analysis and 17% decline curve, InChianti had 29% survival and 21% decline curves. LASA had 41% survival analysis and only 14% decline curves. EPESE had a small number of publications with survival analysis, 18%, or decline analysis, 17%. BLSA had 33% of publications with decline curves but no publication with survival analysis. KLoSHA, which is a relatively young study, had no decline curve or survival analysis publications at this point. The window of publications we have sampled is too narrow to allow the correct perspective on studies' type of publications because this is related to the study evolution; however, the selected publications do provide a reliable snapshot of missing data current practices within the publications category.

DISCUSSION

In the past decade, with the increasing rate of older people, major efforts are being invested in long-term longitudinal population research in aging. The general aim of these studies is to improve our understanding of age-related functional and health changes and to identify factors that may promote successful aging. These objectives are being fulfilled in various aspects; however, with the prolongation of studies' follow up, missing data and attrition are growing, and selection bias limits the conclusions drawn from study findings. As a result, despite the fact that attrition and missing data are really a nuisance, they cannot be ignored. In this article we wanted to obtain a snapshot of the presence of missing data methods that are used in the latest geriatric publications, in the light of the recent recommendations on methods for handling missing data by an expert panel, based on seminal statistical developments of the past three decades [1].

To this end, we reviewed a sample of 103 scientific articles from six key studies of older persons who are living independently in their respective communities. The majority of the publications, 77%, had a detailed account of data selected for analysis but only 28% employed some form of missing data methodology that allows for all available cases in the analysis. Only a few studies (<10%) employed simple methods dedicated to missing data such as comparison of selected and omitted datasets and single imputations for baseline

missing values. Multiple imputations (MI) was conducted only in handful of studies (less than 5%). None of the studies reported the use of advanced methods such as Inverse Probability Weighting (IPW) regression or Joint Models (JM) to account for non-random incomplete data. Although studies varied in terms of geography and biomedical interest, they regarded missing data in a similar fashion.

Studies with long follow ups, such as the Health ABC and LASA, employed longitudinal analyses to describe decline and survival curves, in comparison to the young KLoSHA study. Ferraro described nearly ten years ago the entrance of longitudinal analysis to publications in the Journals of Gerontology [4]. He showed that since the journal's inception in 1945, cross-sectional analyses were most common and that longitudinal analyses entered the scene in mid-1995, reaching about 45% usage of longitudinal analysis. He noted that one of the major problems scientists are facing with the analysis of longitudinal data is attrition, and that prolonged follow up requires from researchers not only application of advanced longitudinal methods but also giving attention to attrition as a source of selection bias.

We are speculating that lack of attention to the attrition bias may diminish or even mask important findings from longitudinal studies in the older population. For example, Yaffe and colleagues [4] studied the effect of β-amyloid on cognitive decline for a subset of n=997 participants out of the n=3075 Health ABC cohort. They noted that at baseline those selected for analysis were more likely to be female, black, and of lower mean education. In a linear mixed model, an overall nine-year decline of about 5 points in the 3MS cognitive score was demonstrated; the difference between the lower and upper tertiles was about 3 points. They also showed that as cognitive reserves plays a modifying role in this association, the decline in score was about twice as large in the below-high-school-diploma compared to above-high-school-diploma subsets. It is possible to speculate that the protective effect associated with cognitive reserve is even larger because: a) the analysis subset had lower education than the complete cohort and b) possible uneven attrition of the sturdy participants, which together perhaps lead to underestimation of the mental reserve heterogeneity, thus limiting its full moderating impact on cognitive decline in the statistical modeling.

Selection bias due to attrition in prolonged geriatric studies may lead to lack of anticipated associations or their attenuation. More study participants at the most fragile stages in life will fail to complete parts of the study forms, and collected information will be over-represented by stronger participants. The problem,

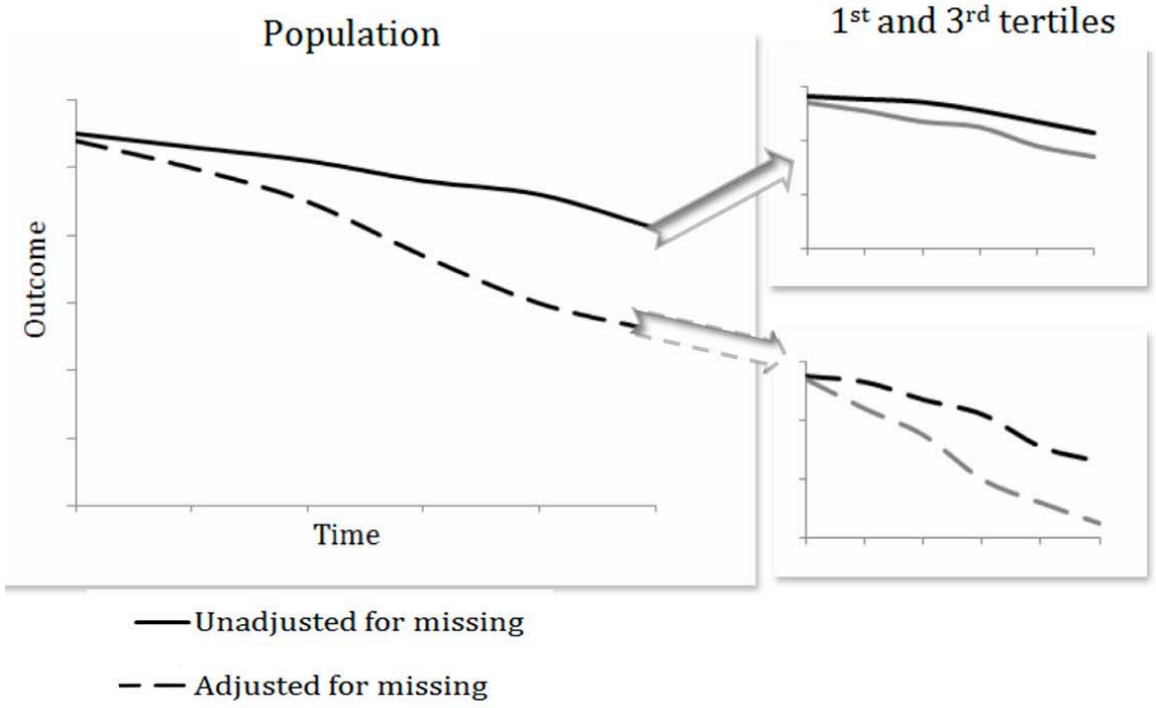


Figure 2: This conceptual figure illustrates that for the entire population, accounting for missing data will result in a steeper decline curve over time, dashed vs. solid lines in the left panel. This in turn leads to more heterogeneity and potentially a larger difference in decline between 1st and 3rd tertiles of an ecological factor, upper vs. lower right panels.

however, has another side; obtaining the imaginary decline curve of the perished participants is unreasonable too. It is therefore required to use in the model all available information until the point of death and to avoid extrapolation beyond that point. Figure **2** provides a conceptual illustration of the potential bias due to attrition in late life. The association between an environmental factor and decline of a score is illustrated both unadjusted and adjusted for missing data in the population (solid vs. dashed line, left panel). Each of the two all-population curves is then split to provide the comparison between 1st and 3rd tertiles, the uncorrected curves (right upper panel) represent more sturdy participants who survived, with smaller tertiles effect, while the corrected curves–(right lower panel) allow for greater heterogeneity by including information about missing data, which in turn leads to curves' further divergence and thus demonstrating larger impact on the decline.

Although this review is limited in scope and number of studies, it provides a valuable snapshot of missing data methodology employed by recent geriatric cohort studies. Our review shows that publications in the past years provide detailed descriptions of the selected analysis dataset, but the majority fails to account for it in the analysis and to show some form of sensitivity analysis.

Pertinent methods described in a recent monograph [1] may be used to meet the challenge of attrition in geriatric studies include Multiple Imputations (MI), Inverse Probability Weighting (IPW), and Joint Modeling (JM). In MI the missing data are imputed by the EM algorithm [5,6] iteratively in E/M steps: E – estimating the missing data based on the M – Maximum likelihood estimates, until convergence. S imputed datasets are generated using Markov Chain Monte Carlo (MCMC) simulation. These are used for obtaining the MI estimates from the pooled S estimates and confidence coverage. MI was first proposed by Rubin in 1978 [7] and it is available in the major statistical packages, SAS PROC MI, SPSS, R, STATA, and other dedicated software. MI is rather simple to implement with the ready-made software, and it improves on the single imputation variance underestimation; however it relies on the Missing-At-Random assumption, which may be challenging in the geriatric depletion setting, and it gives no concrete parameterization of the missing process.

IPW for simple mean estimation requires a weighted mean; the weights are based on the probability of being missing, given the available data as obtained with logistic regression. In the longitudinal setting, this approach is extended using the Generalized Estimating Equations (GEE) [8] method to first obtain the serial missing probabilities and then the weighted regression estimates. Hogan [9] provides an example and SAS code for this method. IPW requires a plain repeated measures design and it can be useful when missing data is assumed monotone and at-random, yet the parameter estimates are sensitive to the assumed missing probabilities. Rotnitzky [10] and Tsiatis [11] proposed the Double Robustness IPW to protect against the possible bias.

Joint Modeling of longitudinal and time-to-event data, as presented by Tsiatis and Davidian [12], is a method for jointly modeling the survival and longitudinal process using the shared-parameter model. This model offers a framework to control for the survival process while studying the decline curve over time and thus appears suitable for geriatric cohort studies. Rizopoulos, in a recent book [13], made this model available for R users by the JM package [14]. This modeling approach is somewhat more complex as it offers to first estimate the mixed model [15, 16] for the longitudinal part, then to obtain the hazard model using Cox proportional hazard model, and finally obtaining the estimation of the joint model including a shared parameter for scaling the association of the two processes. This is basically an MLE method that does not require monotone missingness or pre-set timing of measurements. The drawback of this model is that it can be computationally intensive, yet the advantage is that it also allows for studying both event and longitudinal processes and to test hypotheses regarding the interaction between the two processes. In contrast to MI and IPW, this model provides a framework for individual dynamic prediction [17, 18] in addition to group parameter and mean predications.

Sensitivity analysis of the results is advised because the true values of the complete data remain unknown and approximated by modeling. Such sensitivity analysis may include evaluation using a different model or different parameterization. In the Joint Modeling framework, for example, different assumptions regarding the nature of the association between the longitudinal and event process may be explored. The JM CRAN procedure [17] offers several association forms including value or lagged value of the fixed or the random effect, and slope association using the derivative of the longitudinal process [13 section 5.1].

We conclude by encouraging researchers to try using this methodology for missing data analysis and to consider some form of sensitivity analysis in order to achieve full and unbiased findings.

REFERENCES

[1] Panel on Handling Missing Data in Clinical Trials. The prevention and treatment of missing data in clinical trials. National Academy Press 2010.

[2] Ahrens W, Pigeot I. Handbook of epidemiology. Springer 2005.

[3] Ferraro KF, Kelley-Moore JA. A half century of longitudinal methods in social gerontology: Evidence of change in the journal. J Gerontol Series B: Psychol Sci Soc Sci 2003; 58(5): S264.

[4] Yaffe K, Weston A, Graff-Radford NR, Satterfield S, Simonsick EM, Younkin SG, *et al*. Association of plasma β-amyloid level and cognitive reserve with subsequent cognitive decline. JAMA 2011; 305(3): 261. http://dx.doi.org/10.1001/jama.2010.1995

[5] Allison PD. Missing data. Thousand Oaks, CA: Sage Publications 2001.

[6] Little R, Rubin D. Statistical analysis with missing data. 2nd edn. Wiley & Sons 2002.

[7] Rubin DB. Multiple imputation for nonresponse in surveys. Wiley & Sons 1987. http://dx.doi.org/10.1002/9780470316696

[8] Zeger SL, Liang KY, Albert PS. Models for longitudinal data: A generalized estimating equation approach. Biometrics 1988; 44(4): 1049-60. http://dx.doi.org/10.2307/2531734

[9] Hogan JW, Roy J, Korkontzelou C. Handling drop-out in longitudinal studies. Stat Med 2004; 23(9): 1455-97. http://dx.doi.org/10.1002/sim.1728

[10] Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with nonignorable nonresponse. J Am Statist Assoc 1998; 93(444): 1321-39. http://dx.doi.org/10.1080/01621459.1998.10473795

[11] Tsiatis A. Semiparametric theory and missing data. Springer 2006.

[12] Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: An overview. Statistica Sinica 2004; 14(3): 809-34.

[13] Rizopoulos D. Joint models for longitudinal and time-to-event data: With applications in R. CRC Press 2012. http://dx.doi.org/10.1201/b12208

[14] Rizopoulos D. JM: An R package for the joint modelling of longitudinal and time-to-event data. J Statist Soft 2010; 35(9): 1-33.

[15] Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. Springer 2009.

[16] Diggle P, Heagerty P, Liang K, Zeger S. Analysis of longitudinal data. Oxford University Press 2013.

[17] Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and Time-to-Event data. Biometrics 2011; 67(3): 819-29. http://dx.doi.org/10.1111/j.1541-0420.2010.01546.x

[18] Proust-Lima C, Taylor JM. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: A joint modeling approach. Biostatistics 2009; 10(3): 535-49. http://dx.doi.org/10.1093/biostatistics/kxp009

Received on 23-09-2013 Accepted on 20-10-2013 Published on 11-11-2013

http://dx.doi.org/10.6000/1929-6029.2013.02.04.5